NASA

SQT

E85-10057

NASA-CR-174229

164000-13-P

Eighth Type II Quarterly Status
and Technical Progress Report

STUDY OF SPECTRAL/RADIOMETRIC CHARACTERISTICS OF THE THEMATIC MAPPER FOR LAND USE APPLICATIONS

21 June 1984 — 20 September 1984

WILLIAM A. MALILA
MICHEAL D. METZLER

OCTOBER 1984

Contract NAS5-27346
NASA Goddard Space Flight Center
Greenbelt Road
Greenbelt, Maryland 20771

ENVIRONMENTAL
RESEARCH INSTITUTE OF MICHIGAN
BOX 8618 • ANN ARBOR • MICHIGAN 48107

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
164000-13-P		

4. Title and Subtitle
Study of Spectral/Radiometric Characteristics of the Thematic Mapper for Land Use Applications

5. Report Date
October 1984

6. Performing Organization Code

7. Author(s)
William A Malila and Michael D. Metzler

8. Performing Organization Report No.
164000-13-P

9. Performing Organization Name and Address
Environmental Research Institute of Michigan
P.O. Box 8618
Ann Arbor, MI 48107

10. Work Unit No.

11. Contract or Grant No.
NAS5-27346

12. Sponsoring Agency Name and Address
National Aeronautics and Space Administration
Goddard Space Flight Center
Greenbelt, MD 20771

13. Type of Report and Period Covered
Quarterly Status and Technical Progress
21 June - 20 Sep 1984

14. Sponsoring Agency Code

15. Supplementary Notes
Mr. Harold Oseroff (Code 902) is serving as Technical Officer and Mssrs. Brian Markham and James Irons (Code 923) are serving as Science Representatives for NASA/GSFC.

16. Abstract

Progress during ERIM's eighth quarter of effort under the Landsat-4 and 5 Image Data Quality Assessment program for the Thematic Mapper is described.

Analyses of Landsat-5 TM radiometric characteristics were performed. Effects which had earlier been found in Landsat-4 TM data were found to be present in Landsat-5 data as well, including:

1) Scan-direction-related signal droop
2) Scan-correlated level shifts
3) Low-frequency coherent noise

Coincident Landsat-4 and 5 raw TM data were analyzed, and band-by-band relationships between the two sensors were derived.

Earlier efforts which developed an information-theoretic measure of multispectral information content were continued, comparing TM and MSS information content.

17. Key Words
Radiometric Calibration
Landsat 4, Landsat 5
Thematic Mapper
Noise
Information Theory

18. Distribution Statement
Initial distribution is listed at the end of this document.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
UNCLASSIFIED	UNCLASSIFIED	33 + iii	

ERIM — INFRARED AND OPTICS DIVISION

Report No. 164000-13-P

Eighth
Type II Quarterly Status and Technical Progress Report
21 June 1984 - 20 September 1984

for

Study of Spectral/Radiometric Characteristics of the Thematic Mapper for Land Use Applications

under

Contract NAS5-27346

with

NASA Goddard Space Flight Center
Greenbelt Road
Greenbelt, Maryland 20771

Submitted by

Environmental Research Institute of Michigan
P.O. Box 8618
Ann Arbor, Michigan 48107

Prepared by:
FOR
William A. Malila
Principal Investigator

Approved by:
Robert Horvath
Manager, Information Processing Department

October 1984

Table of Contents

1. OBJECTIVE 1

2. TASKS 1

3. STATUS AND TECHNICAL PROGRESS 1
   3.1 PROBLEMS 2
   3.2 ACCOMPLISHMENTS 2
      3.2.1 Landsat-5 TM Noise and Droop Effects 2
      3.2.2 TM Landsat-4 vs Landsat-5 Radiometric Comparison 4
      3.2.3 Information-Theoretic Comparison of TM and MSS Data 4
   3.3 SIGNIFICANT RESULTS 6
   3.4 PUBLICATIONS AND PRESENTATIONS 7
   3.5 RECOMMENDATIONS 7
   3.6 FUNDS EXPENDED 7
   3.7 DATA RECEIPTS 7

APPENDIX A 20

DISTRIBUTION LIST 33

Eighth Quarterly Report

STUDY OF SPECTRAL/RADIOMETRIC CHARACTERISTICS OF THE THEMATIC MAPPER FOR LAND USE APPLICATIONS

1. OBJECTIVE

The objective of this investigation is to quantify the performance of the TM as manifested by the quality of its image data in order to suggest improvements in data production and to assess the effects of the data quality on its utility for land resources applications. Three categories of this analysis are: a) radiometric effects, b) spatial effects, and c) geometric effects, with emphasis on radiometric effects.

2. TASKS

Four tasks have been established to address the above objective. The first three are to study radiometric performance, spatial performance, and geometric performance, respectively, while the fourth is to study spectral characteristics. In keeping with the identified objective, the radiometric performance study is our major task.

3. STATUS AND TECHNICAL PROGRESS

During this eighth quarterly reporting period, a more in-depth analysis of Landsat-5 TM radiometric characteristics was performed. Scan-direction-related 'droop' effects and scan-related level shifts were found and examined using both nighttime data and a relatively uniform scene of daytime data. Coincident Landsat-4 and Landsat-5 TM data were compared, and band-by-band correlations were established for the values prior to radiometric correction. Earlier efforts which developed an information-theoretic measure of information content in multispectral data were continued and extended.

3.1 PROBLEMS

Non-receipt of radiometrically and geometrically corrected data from the coincident-coverage scene of Landsats-4 and 5 has precluded a complete comparison of TM data from the two satellites.

Definitive quantitative analyses of noise effects were hampered by the fact that the night scene (5-0052-02182) data we received were incomplete. In Quadrant 3, Lines 36-52 were missing from Band 1, as was Line 102 from Band 1, Quadrant 4.

3.2 ACCOMPLISHMENTS

Accomplishments in three technical areas are described below.

3.2.1 Landsat-5 TM Noise and Droop Effects

Earlier efforts by the authors[1,2,4,5,9] resulted in characterization of noise effects in Landsat-4 TM which had not previously been reported. Three types of noise were reported:

1. An effect related to scan-direction was discovered, whereby the mean signal level (in reflective bands) was observed to decay ('droop') as the active scan progressed during daytime scenes, and similarly to rise as a function of time in nighttime scenes[2,4].

2. A shift of the mean signal of several adjacent detectors in Band 1 upward or downward for one or more scans was first reported by Kieffer[3]. We found that the effect was not limited to Band 1, and provided the initial characterization of this effect[4,5]. This included discovery that all level shifts were strictly correlated among the affected detectors, with two basic level shift patterns being present in all detectors (Bands 1-5,7) to varying degrees. Odd and even detectors were generally 180° out of phase with one another, one set shifting up when the other set shifted down. The two patterns were characterized by Band 1 Detector 4 and Band 7 Detector 7, respectively.

   This effect has been examined by several other investigators[e.g., 3,7-9]. Correction mechanisms have been proposed[4-9], and in the case of the Canada Centre for Remote Sensing (CCRS), implemented in an operational radiometric correction processing system[7].

3. A low frequency (approximately 400Hz or 262-264 pixel wavelength) coherent noise was detected in Bands 1-5 and 7[9]. This low amplitude (<0.75DN) effect was observed in nighttime data only.

The tools developed for Landsat-4 TM noise analysis were applied to Landsat-5 TM data. Those analyses revealed artifacts in Landsat-5 TM data which correspond to the three types of noise described above. Only raw, uncorrected data

were available for analysis, but the effects are expected to be found in corrected data as well, if they follow the pattern we found in Landsat-4 TM data. Descriptions of those findings are provided below; in-depth analysis is planned for the next reporting period.

Within–Line Droop. Both nighttime reflective band data and data from relatively homogeneous daytime scenes were used to analyze Landsat-5 TM data for the presence of the within-line droop effect discovered in Landsat-4 TM. Scene 5-0052-02182 (Harrisburg, PA) provided the nighttime data, while the daytime data were from Scene 5-0014-15460 (Alabama). The daytime scene, although not ideal in terms of spatial and spectral homogeneity, had the advantage of being coincident with Landsat-4 TM Scene 4-0608-15463, thus allowing direct comparison of effects in the two sensors.

The data examined revealed a 'droop/rise' effect in Landsat-5 TM data which appeared nearly identical to that observed with Landsat-4. The magnitude of the Landsat-5 effect appears to be somewhat less than for Landsat-4 TM, but the direction (nighttime 'rise' and daytime 'droop'), and time constants (approximately 800-1000 pixels) appear very similar. Figure 1 illustrates this effect for Band 4 of both Landsat-4 and Landsat-5 Thematic Mappers. Apparently the causal mechanism was not removed by the modifications made to Landsat-5 TM prior to launch.

Scan–Correlated Level Shifts. Scene 5-0052-02182 provided the radiometrically uniform scene data essential for optimal extraction of scan-correlated level shifts. Unfortunately, the missing data problems mentioned above with regard to this particular scene hampered the analysis of the level shift effect also.

Figure 2 illustrates the scan-line mean signal returned by each of the detectors in Band 3. The level shifts observed in the Landsat-5 TM data appear to fit a single pattern as opposed to the two patterns identified for Landsat-4 TM. The Landsat-5 level shifts may be characterized by the detectors in Band 3, and for this reason are identified as Type 5-3 level shifts by Barker[8]. Although all Band 3 detectors exhibit this level shift behavior, it may be found in nearly all the detectors in all the reflective bands to some extent. As with the level shifts found in Landsat-4 TM data, some of the detectors shift with a phase directly opposed to the phase of the prototype (Band 3) detectors. These phase relationships are seen most clearly in Bands 1, 5, and 7, where, in general, the odd numbered detectors shift in phase with the Band 3 detectors, and the even-numbered detectors have a level shift pattern 180° out of phase with Band 3 shifts.

Low–Frequency Coherent Noise. Low-frequency (approximately 400Hz) coherent noise was observed in Landsat-4 TM nighttime reflective data. This noise was seen to be of low amplitude and present in all non-thermal bands. Preliminary analysis of the one nighttime Landsat-5 TM scene we have indicates that this noise is present in Landsat-5 TM data as well. Figure 1 illustrates this noise (along with the nighttime 'rise') for Band 4 of both sensors. The approximately 260-pixel period is clearly present in these data, even though the Landsat-5 data had only minimal filtering applied. The low amplitude (<0.02DN) of this noise would prevent it from being observed in daytime data if the effect were additive. Previous

examination of daytime Landsat-4 TM data did not reveal any discernable effects of this coherent noise.

3.2.2 TM Landsat-4 vs Landsat-5 Radiometric Comparison

As Landsat-5 was moving to its WRS orbit after launch, an opportunity was present to acquire near-simultaneous Landsat-4 and 5 TM data over Alabama. Analysis of the raw data can provide an indication of the radiometric consistency and linearity of the two sensors; analysis of the radiometrically corrected data provides the information necessary to use the two sensors together or interchangeably. Our current analyses were restricted to the raw, radiometrically uncorrected data only, because the corrected data tapes have not been received.

Approximately 30 regions were selected from Scene 4-0608-15463 which spanned the scene dynamic range in each of the seven spectral bands. These regions were then extracted from the Landsat-5 data (Scene 5-0014-15460), and region mean signal values were calculated for each band of each sensor. Band-by-band plots of these data revealed a high degree of linearity and correlation, also indicated by R^2 values of >0.995 in all cases. The relationships between Landsat-4 and Landsat-5 TM data over the dynamic range present in these scenes are illustrated by Figures 3(a)-(g) for Bands 1-7, respectively. In general, the gains of Landsat-5 TM Primary Focal Plane bands (1-4) are slightly greater than for Landsat-4, and the Landsat-5 Cold Focal Plane bands (5-7) have lower gains. Band 6 of Landsat-5 has a much lower gain than Landsat-4 in these uncorrected data, possibly due to different shutter reference temperatures in the two sensors.

The actual regression coefficients have limited value since these relationships were developed for radiometrically *uncorrected* data. The radiometric correction process, in providing the conversion of signal counts (DN) to radiance, would presumably remove the differences between the two sensors, i.e., the regressions would each have unity gain and zero offset. Examination of radiometrically corrected data will demonstrate the success of the process in achieving that goal.

3.2.3 Information-Theoretic Comparison of TM and MSS Data

In the sixth quarterly report on this contract[11], an information-theoretic measure of the information content of multispectral scanner data was derived and applied to Landsat TM and MSS data. The effort was continued during the current quarter, with some additional analysis results. The additional analyses are summarized in this section. Appendix A contains a paper, describing the overall effort, which was prepared for presentation at the Eighteenth International Symposium on Remote Sensing of Environment and for publication in the proceedings of that symposium. Simple numerical examples to help one understand entropy concepts were generated and are in the Appendix, but will not be repeated here.

Data–Space Descriptions. In the previous report, graphs were presented of the quantities that compared information capacities and data-set characteristics. The

diagram in Figure 4 helps describe the various terms used to designate spectral data-space characteristics, while Table 1 quantifies observed values for several cases that were considered. First, the system-design capacities of the Landsat-4 TM and MSS are presented, in terms of the number of bits transmitted to the ground and/or recorded on computer-compatible tapes (CCTs). The system design capacity is the sum of the available bits in the various spectral bands (equal to the logarithm to the base two of the product of the maximum number of digital levels in those bands, hence a "volume".) For TM, the number of bits recorded on CCTs is the same as that transmitted (8 bits/channel). For MSS, however, the six-bit telemetered data are expanded to seven bits on the CCTs, with only an apparent gain of information. Comparisons involving MSS are given for both seven-bit data (since that is the form in which we received them) and data after a degradation to six bits was performed. The greater information potential of the TM system design (reflective bands), as compared to the MSS system, is quantified as 48 vs 24 bits in telemetered data.

Figure 4 also portrays the "hypercube" volume or data-space volume spanned by multispectral data. These volumes are computed by summing the bit equivalents of the observed data-value ranges (*max − min + 1*) in each band being considered. Upon comparing the fractions of their total data-space volumes that are spanned by data from the agricultural scene, one observes that the TM data fall nine bits short of capacity, while the MSS data fall approximately six bits short of capacity.

Actual data dispersion volumes (see Figure 4 and Table 1) were found to be substantially smaller than the "hypercube" volumes. Results for both real and synthetic data shown in Table 1 represent the actual information content of the data. (Note that these values for actual information are substantially smaller than reported elsewhere for similar comparisons in which the "hypercube" volumes are treated as the information content[12].) The number of observations analyzed established a maximum limit on each entropy value. The concentration of multiple observations (pixels) into individual spectral cells reduces the information content below the potential maximum. The data sets described in Table 1 show very little tendency for TM pixels to cluster, due to the very large system capacity, spectral diversity, and fine gradation of the TM bands. The MSS data show definite tendencies for multiple observations in spectral cells.

Table 1 also shows that the TM data represent 3.3 bits more information than the MSS sensor data, with approximately two bits being associated with spatial resolution (pixel size and number) and the remainder with spectral bands and radiometric resolution. Since the synthetic data have the same number of observations for both TM and MSS, they can be considered to have equal spatial resolutions. Thus, the 2.2-bit difference must be due to their spectral and radiometric properties.

Noise. Noise in multispectral data was not considered explicitly in the results presented herein. Sensor noise effects certainly were present in the real Landsat data and natural variations of crop observations were present in both real and synthetic data. Noise can add variance to signals and increase the number of spectral cells occupied (above that for no noise), thereby creating an apparent information content greater than the true information content of ideal, noiseless

signals. One might address such effects by applying an appropriate quantization factor (greater than unity) to each band to reduce the number of discrete levels present in data sets, and computing the reduced information content.

Summary Discussion. An information-theoretic measure has been defined and applied to Landsat multispectral data, both real and synthesized. Examples of the basic concepts also were generated. The measure does quantify signal dispersion patterns, independently of class membership and distributional assumptions. It also provides an alternate method of measuring the extent to which subsets of bands or transformed variables represent the total pattern. In planning analyses and interpreting results, however, analysts should insure that data sets being analyzed are representative of the problems under consideration.

A number of observations were made from this initial study. The system-design information capacity of TM is much greater than that of MSS. The potential information capacities and the signal "hypercube" volumes of agricultural data were much larger than the information actually represented by signal dispersion patterns in the sets of data values analyzed. For an agricultural data set, the gain in information content of TM over MSS was 3.3 bits, far less than the difference in design capacities. Tasseled Cap transformations preserved the information in original bands and offered a modest savings in bits over those original bands, a fact which might be useful in data compression approaches. There were extremely few multiple occurrences of spectral observations in the TM data set, but a reasonably high number in the MSS data, another indication of TM's finer partitioning of spectral space. For the "best" combinations of variables, entropy magnitudes were more a function of the number of variables than of the type of variables (original bands or transformed). TM had greater entropy values for Brightness variables and Brightness-Greenness pairs than did MSS. Information in the Tasseled Cap Third Component of TM was much greater than that of MSS, both by itself and in combination with Brightness or Greenness, confirming TM's greater dimensionality.

In future studies, it is recommended that additional data sets be analyzed, both with larger sample sizes and with varied scene content; effects of other transformations might also be examined. Noise effects should be investigated through use of quantization factors to degrade radiometric resolutions. It may also be fruitful to investigate approaches to incorporate class membership into information-theoretic measures of multispectral information content.

3.3 SIGNIFICANT RESULTS

(1) Three noise effects present in Landsat-4 TM data were also found in Landsat-5 TM data.

(2) Scan lines were missing from one raw data tape (unit RLUT); if the same effect is present in other tapes, it could cause problems for investigators who are not aware of it.

(3) Radiometric comparisons were established between raw TM data from coincident scenes of Landsat-4 and Landsat-5.

(4) Additional information-theoretic comparisons of Landsat TM and MSS data were made.

3.4 PUBLICATIONS AND PRESENTATIONS

A paper, entitled "Information Theoretic Comparison of Original and Transformed Data from Landsat MSS and TM", by William A. Malila, was prepared for presentation at the Eighteenth International Symposium on Remote Sensing of Environment, Paris, France, October 1984. It will be published in the symposium proceedings. A preprint is included as Appendix A.

3.5 RECOMMENDATIONS

No additional major recommendations beyond those made in previous reports are identified at this time.

3.6 FUNDS EXPENDED

A total of approximately $39,000 was expended during the three months June through August 1984. An amendment to the contract was received to support additional analyses of Landsat-5 data. The cumulative spending through August represents approximately 65% of the amended contract total. Expenditures during the period 1-20 September 1984 are not included in this percentage value.

3.7 DATA RECEIPTS

Raw data tapes (unity RLUT CCT-AT) and calibration data tapes (CALDUMP) were recevied during this quarter for the following scenes:

Alabama	P20/R37	5-0014-15460
Alabama	P20/R37	4-0608-15463
SE Alabama	P20/R38	5-0014-15463
Harrisburg (Night)	P111/R212	5-0052-02182
White Sands	P33/R37	5-0129-17075

Note: Scene 5-0052-02182 was missing lines 36-52 from Quadrant 3 Band 1, and line 102 from Quadrant 4, Band 1 on the Unity RLUT CCT-AT.

Fully corrected data (CCT-PT) were received for two scenes:

NE Alabama	P20/R36	5-0014-15454
Alabama	P20/R37	4-0608-15463

REFERENCES

1. Malila, W.A., M.D. Metzler, D.P. Rice, and E.P. Crist, "Characterization of LANDSAT-4 MSS and TM Digital Image Data", *IEEE Transactions on Geoscience and Remote Sensing*, pp. 177-191, May 1984.

2. Malila, W.A., and M.D. Metzler, *First Quarterly Progress Report on Study of Spectral Radiometric Characteristics of the Thematic Mapper for Land Use Applications*, ERIM Report No. 164000-1-T, December 1982

3. Kieffer, H.H., E. Eliason, P. Chavez, R. Batson, and W. Borgeson, presentation at the Second Landsat-4 Investigations Workshop, NASA Goddard Space Flight Center, January 1983.

4. Malila, W.A., and M.D. Metzler, *Second Quarterly Progress Report on Study of Spectral Radiometric Characteristics of the Thematic Mapper for Land Use Applications*, ERIM Report No. 164000-2-T, March 1983

5. Malila, W.A., M.D. Metzler, and E.P. Crist, *Fourth Quarterly Progress Report on Study of Spectral Radiometric Characteristics of the Thematic Mapper for Land Use Applications*, ERIM Report No. 164000-4-T, September 1983

6. Kogut, J., E. Larduinat, and M. Fitzgerald, "An Analysis of New Techniques for Radiometric Correction of LANDSAT-4 Thematic Mapper Images", Research and Data Systems, Inc., Lanham, MD, October 1983.

7. Murphy, J.M., T. Butlin, P.F. Duff, and A.J. Fitzgerald, "Revised Radiometric Calibration Technique for LANDSAT-4 Thematic Mapper Data", *IEEE Transactions on Geoscience and Remote Sensing*, pp. 243-251, May 1984.

8. Fischel, D., "Validation of the Thematic Mapper Radiometric and Geometric Correction Algorithms", *IEEE Transactions on Geoscience and Remote Sensing*, pp. 237-242, May 1984.

9. Barker, J.L., "Relative Radiometric Calibration of Landsat TM Reflective Bands", *LANDSAT−4 Science Investigations Summary*, vol. I, pp. 140-180, July 1984.

10. Malila, W.A., and M.D. Metzler, *Seventh Quarterly Progress Report on Study of Spectral Radiometric Characteristics of the Thematic Mapper for Land Use Applications*, ERIM Report No. 164000-11-P, August 1984.

11. Malila, W.A., *Sixth Quarterly Progress Report on Study of Spectral/Radiometric Characteristics of the Thematic Mapper for Land Use Applications*, ERIM Report No. 164000-10-P, April 1984.

12. Price, J.C., "Comparison of the Information Content of Data from the LANDSAT-4 Thematic Mapper and Multispectral Scanner", *IEEE Transactions on Geoscience and Remote Sensing*, pp. 272-281, May 1984.

LANDSAT 4 LANDSAT 5

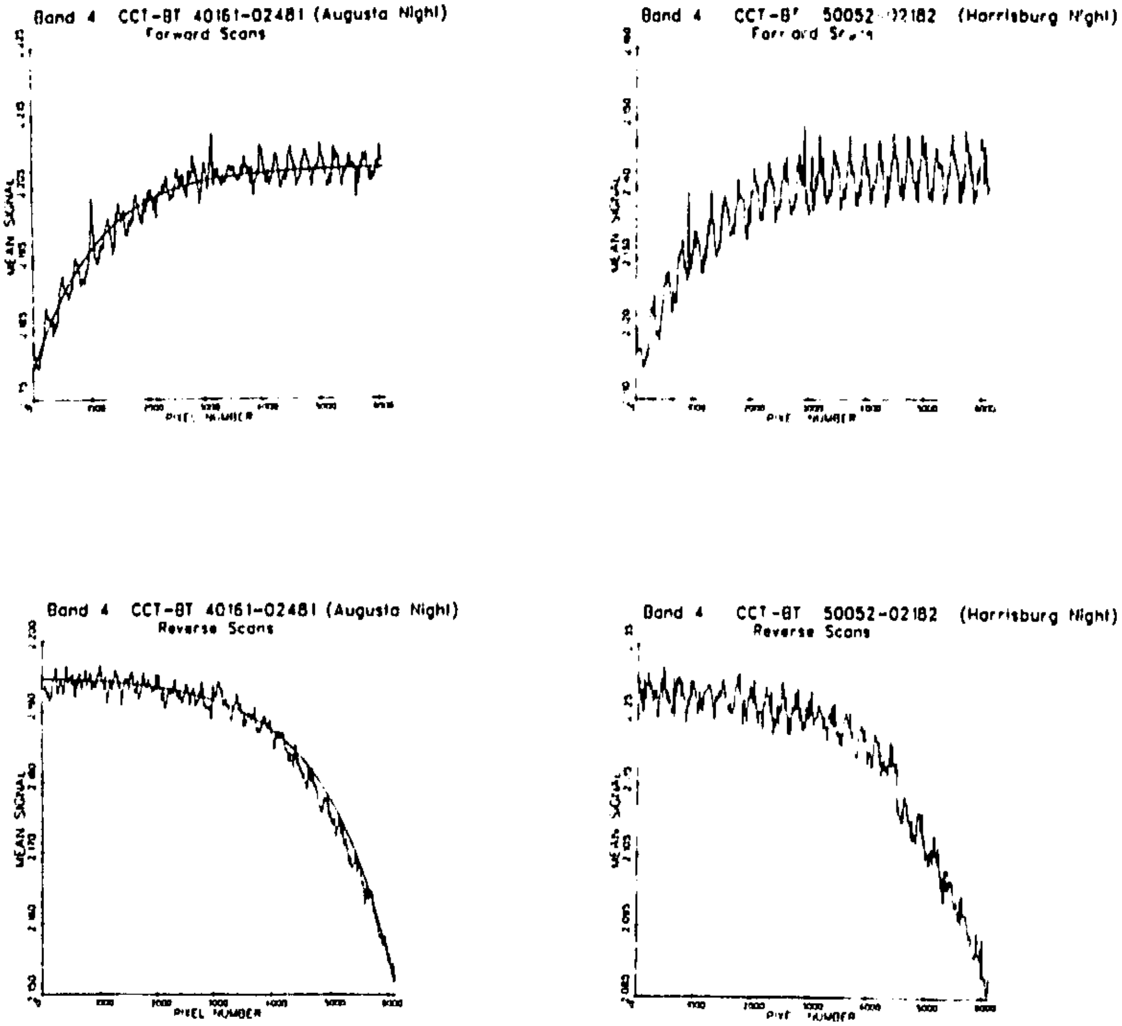

Figure 1. LANDSATS-4 AND 5 NIGHTTIME DROOP EFFECT – BAND 4

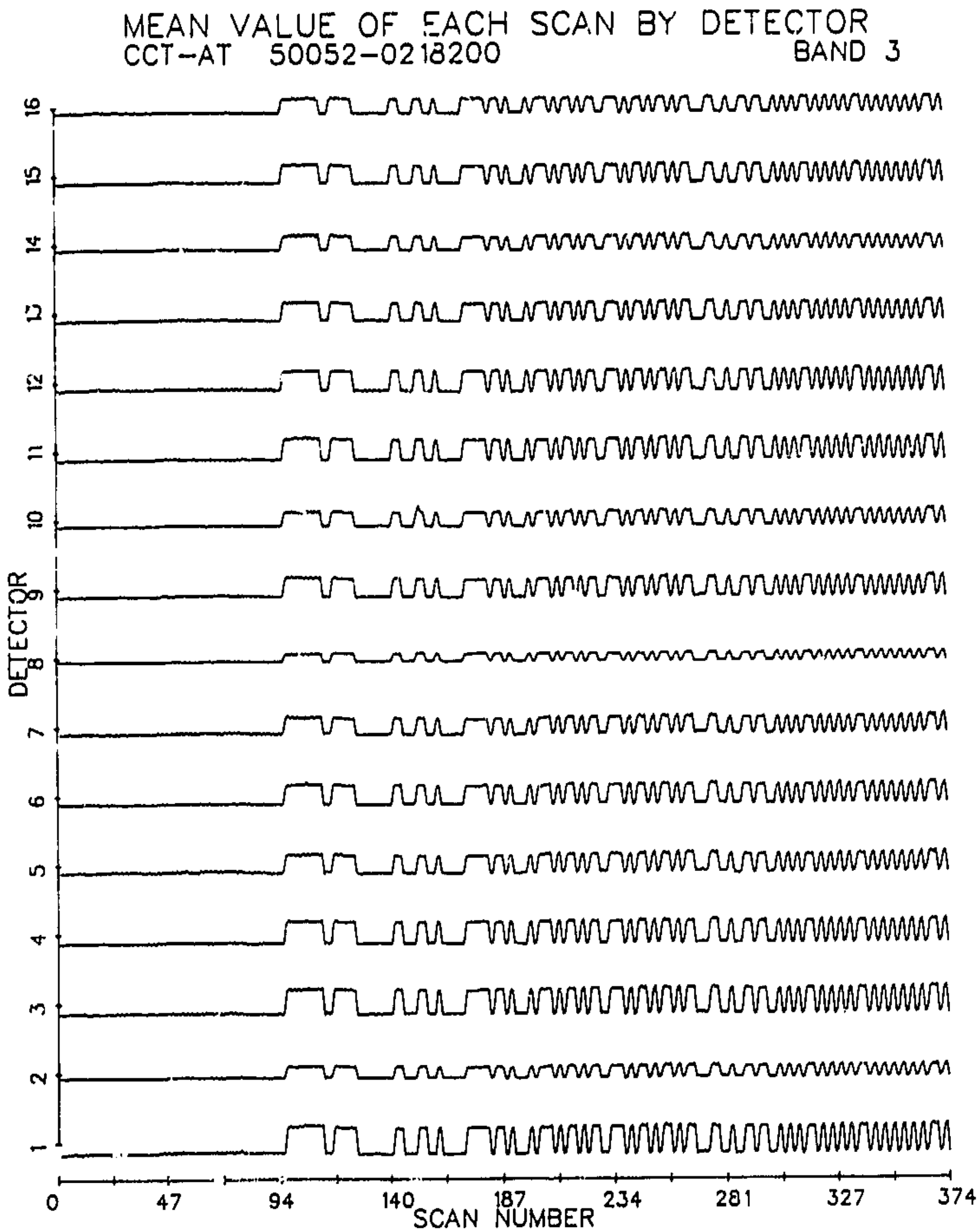

Figure 2. SCAN AVERAGES OF NIGHTTIME SCENE DATA – BAND 3

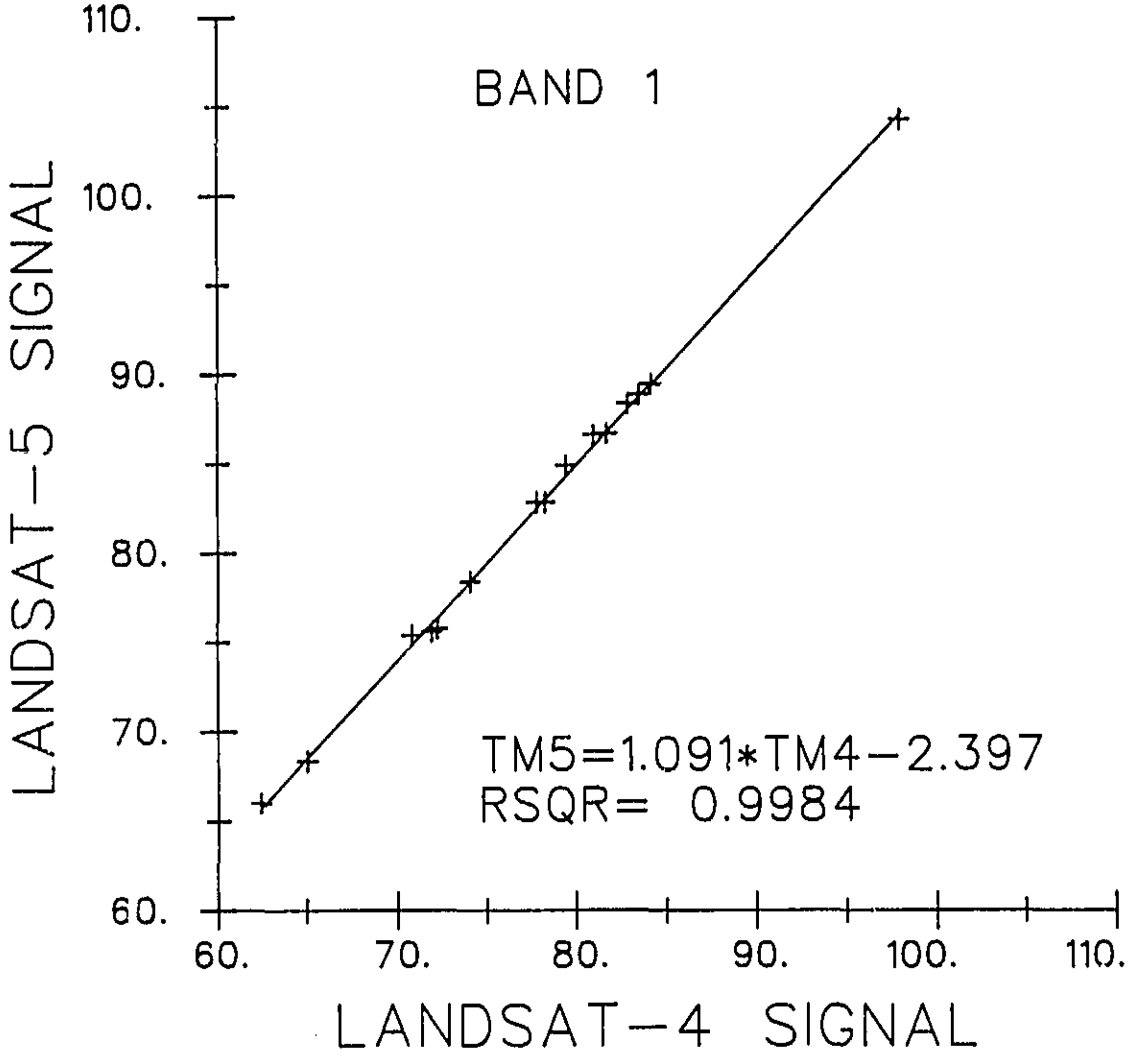

Figure 3(a). RELATIONSHIP BETWEEN LANDSATS-4 AND 5 TM DATA — BAND 1

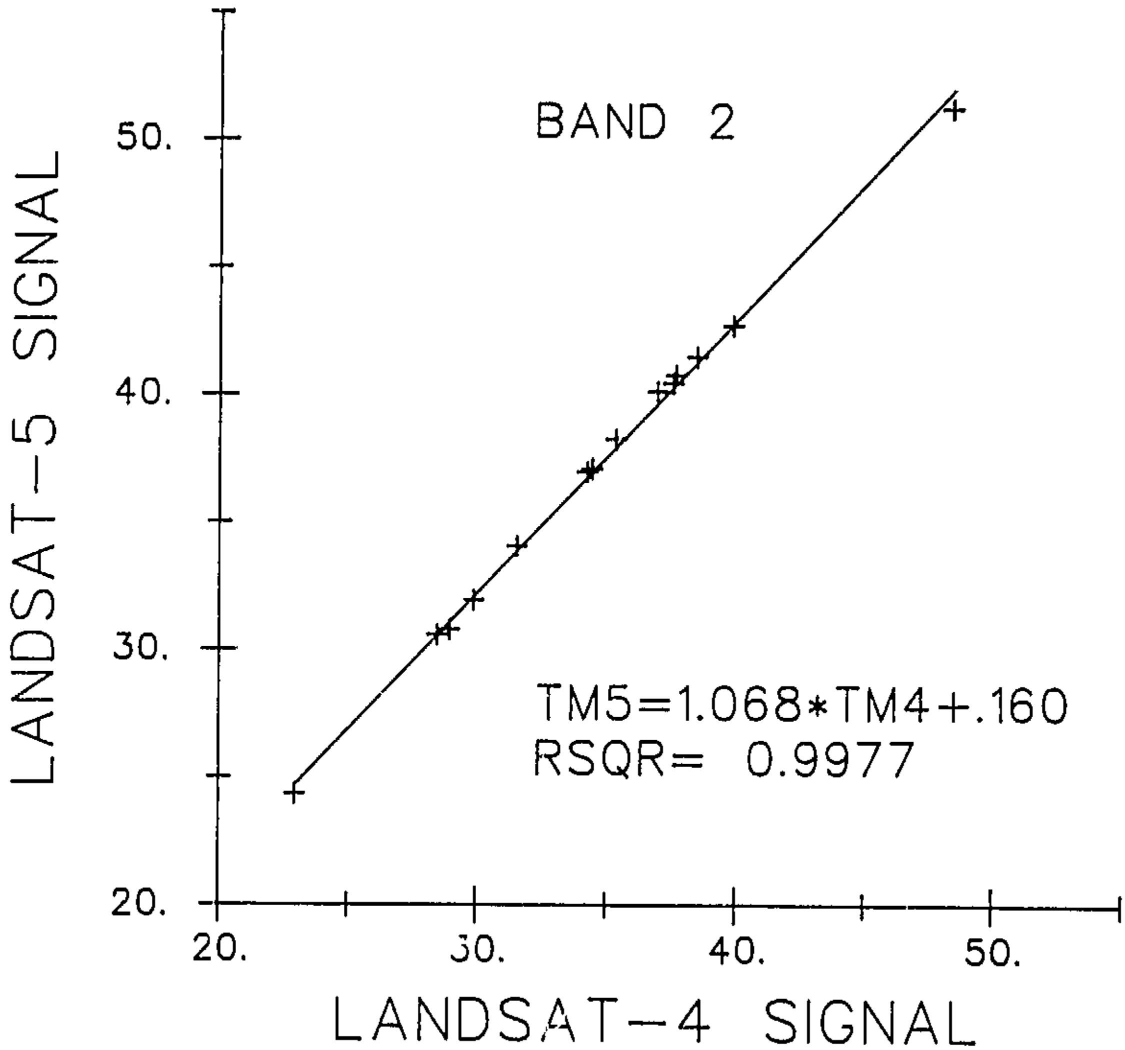

Figure 3(b). RELATION... BETWEEN LANDSATS-4 AND 5 TM DATA — BAND 2

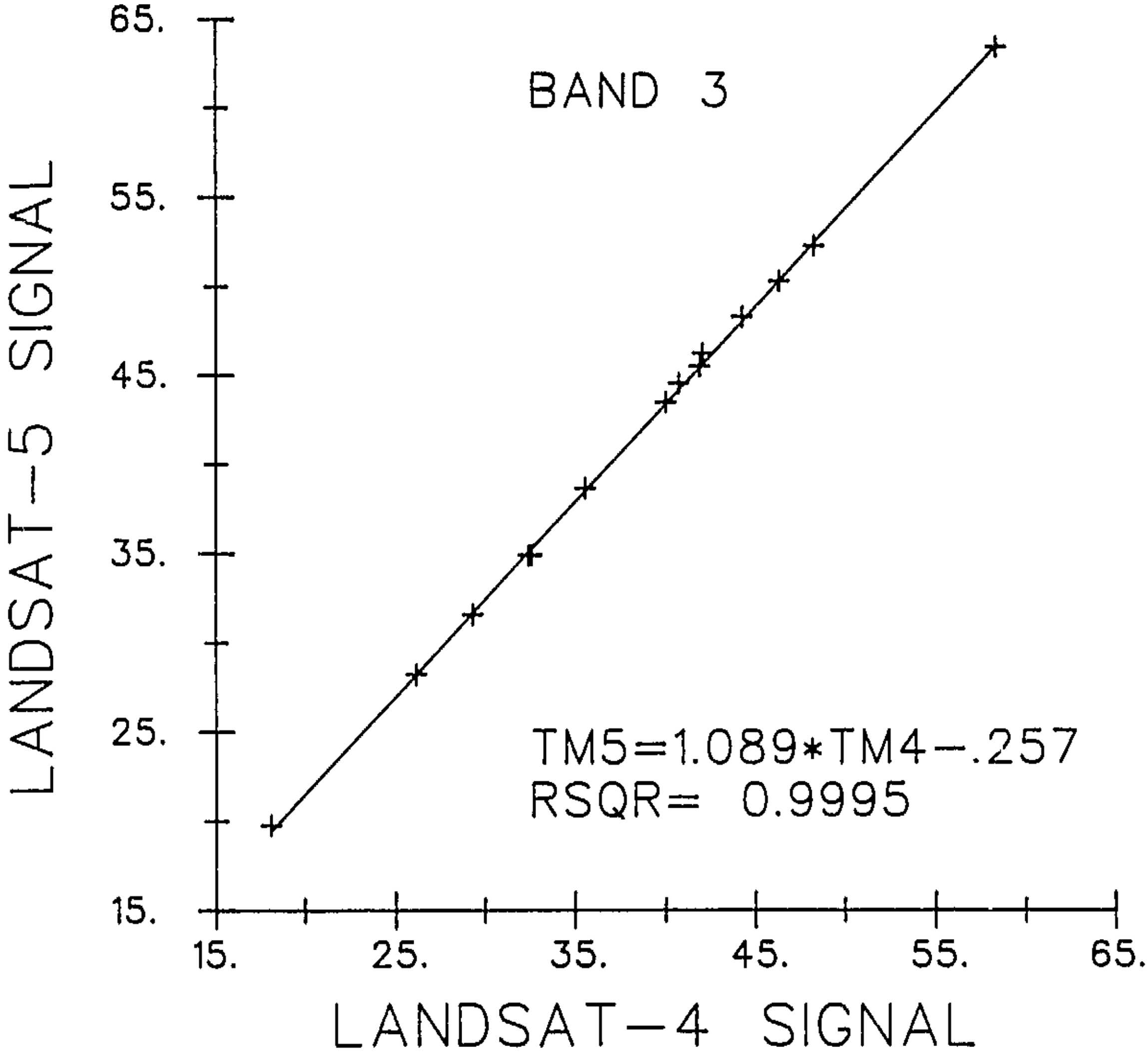

Figure 3(c). RELATIONSHIP BETWEEN LANDSATS-4 AND 5 TM DATA – BAND 3

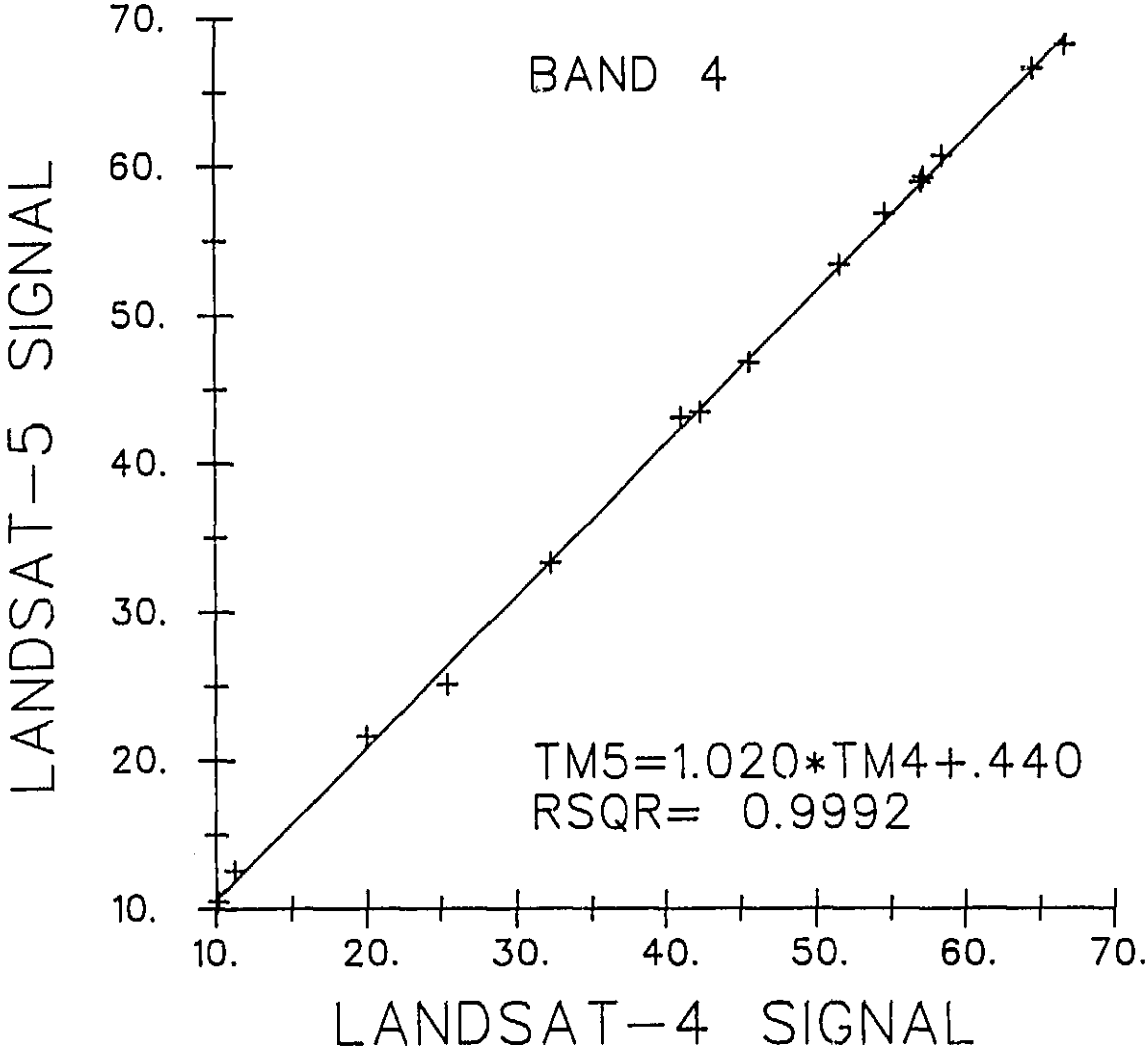

Figure 3(d). RELATIONSHIP BETWEEN LANDSATS-4 AND 5 TM DATA — BAND 4

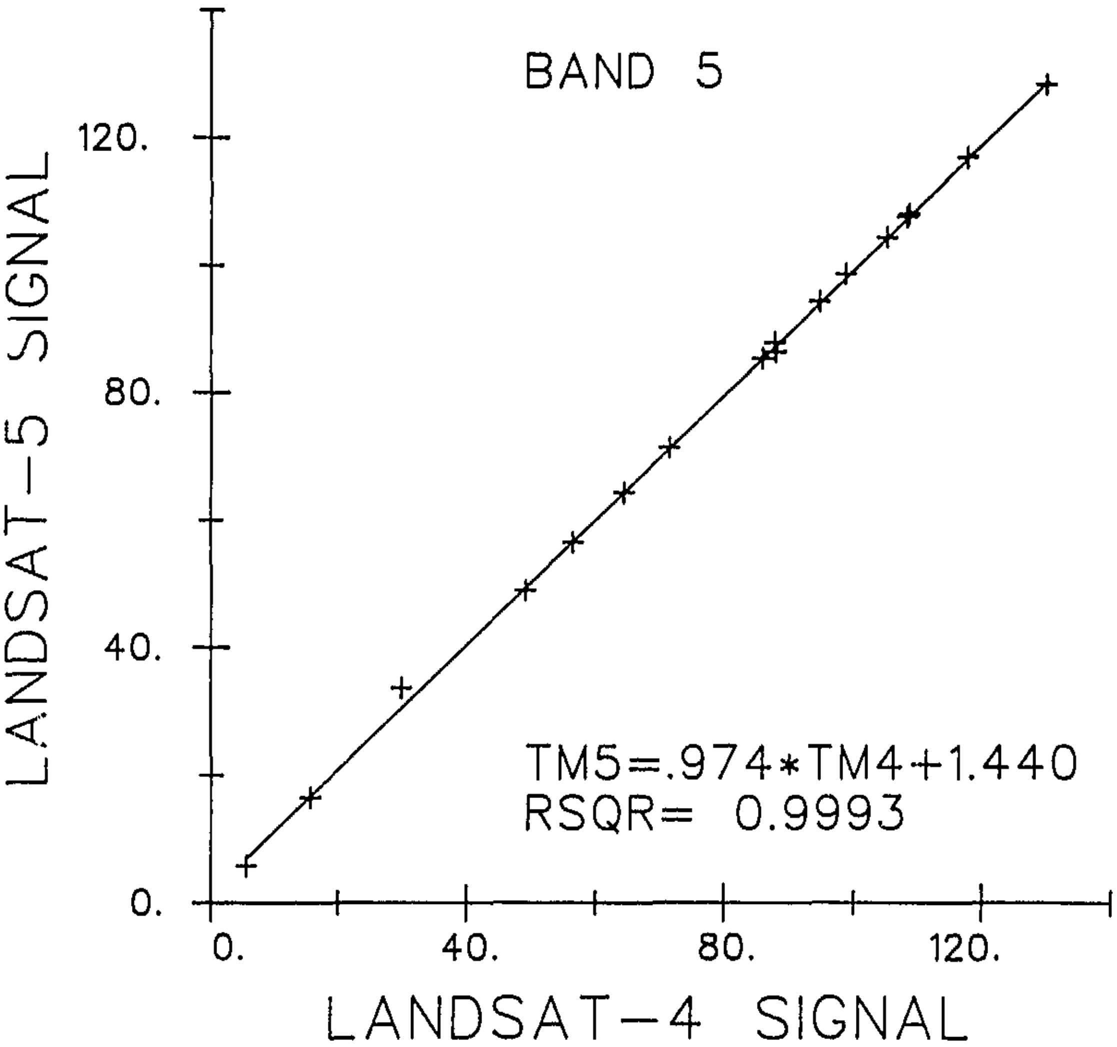

Figure 3(e). RELATIONSHIP BETWEEN LANDSATS-4 AND 5 TM DATA – BAND 5

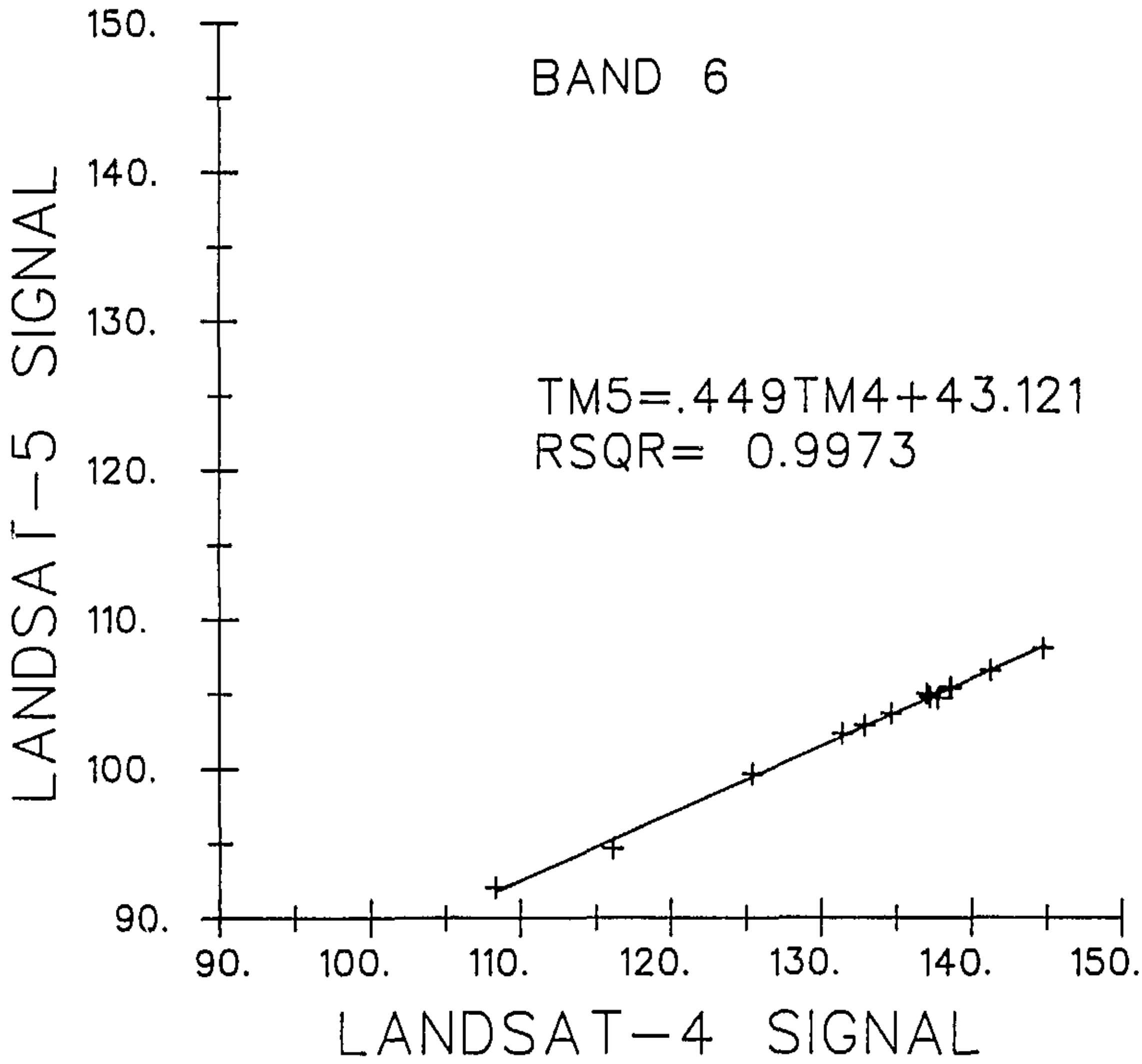

Figure 3(f). RELATIONSHIP BETWEEN LANDSATS-4 AND 5 TM DATA — BAND 6

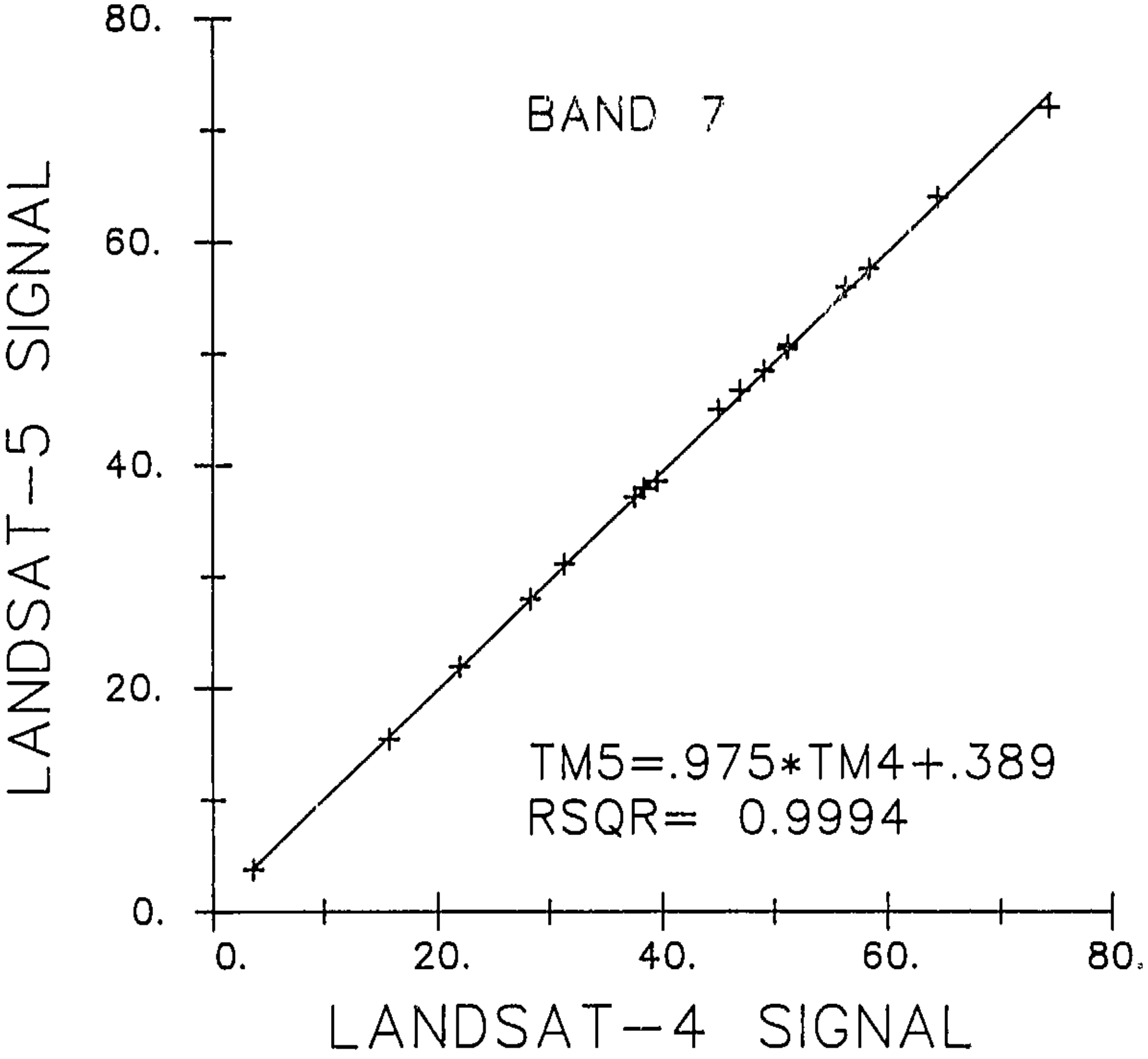

Figure 3(g). RELATIONSHIP BETWEEN LANDSATS-4 AND 5 TM DATA – BAND 7

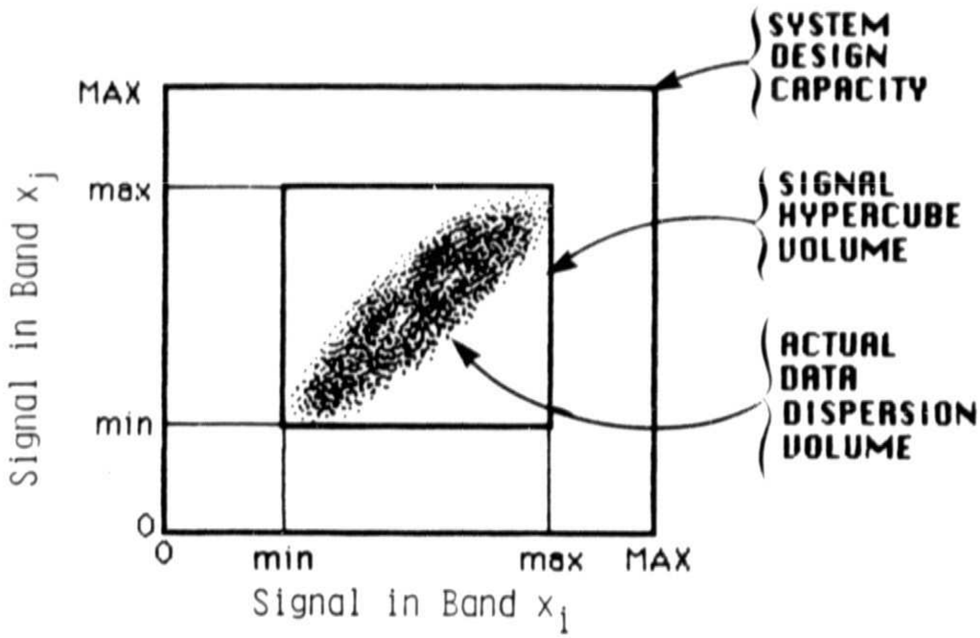

Figure 4. ILLUSTRATION OF VARIOUS MULTISPECTRAL DATA VOLUMES

Table 1. Information Comparison for MSS and Six-Band TM Data Sets

A. VALUES FOR REAL AGRICULTURAL DATA (N. CAROLINA SCENE)

		MSS		SIX-BAND TM		TM GAIN
		NUMBER	BITS	NUMBER	BITS	(BITS)
SYSTEM CAPACITY:	SENSOR	0.17×10^{8}	24	0.28×10^{15}	48	24
	CCT	0.27×10^{9}	28	0.28×10^{15}	48	20
HYPERCUBE VOL.:	SENSOR	0.44×10^{6}	18.7	0.43×10^{12}	38.6	
	CCT	0.32×10^{7}	21.6	0.43×10^{12}	38.6	
DATA DISPERSION PATTERN:						
	· # Observations, H_{max}	3,468	11.8	13,015	13.7	1.91 (Spatial)
MSS CCT: 7 Bits per Band	· # Unique cells	2,898	-	12,903	-	
	· Entropy, H	-	11.4	-	13.7	2.27 (Total)
	· Loss due to spectro-radiometric concentration	-	0.38	-	0.02	0.36 (Spec/Radiom)
MSS Sensor: 6 Bits per Band	· # Unique cells	1,730	-	12,903	-	
	· Entropy, H	-	10.3	-	13.7	3.34 (Total)
	· Loss due to spectro-radiometric concentration	-	1.45	-	0.02	1.43 (Spec/Radiom)

B. VALUES FOR SYNTHETIC AGRICULTURAL DATA
(Assumes equal spatial resolution)

	MSS		SIX-BAND TM		TM GAIN
	NUMBER	BITS	NUMBER	BITS	(BITS)
"SYSTEM" CAPACITY (MSS: 6 Bits/Band)	0.17×10^{8}	24	0.28×10^{15}	48	24
OBSERVED HYPERCUBE VOLUME	0.10×10^{7}	20	0.99×10^{12}	40	20
DATA DISPERSION PATTERN:					
· # Observations; H_{max}	2,276	11.15	2,276	11.15	
· # Unique cells	817	-	2,260	-	
· Entropy, H	-	8.94	-	11.14	2.20
· Loss due to spectro-radiometric concentration	-	2.21	-	0.014	(Spec/Radiom)*

*(TM gain over seven-bit MSS data was one bit.)

APPENDIX A

INFORMATION THEORETIC COMPARISONS OF ORIGINAL AND TRANSFORMED DATA FROM LANDSAT MSS AND TM

William A. Malila

Presented at the Eighteenth International Symposium on Remote Sensing of Environment, Paris, France, October 1-5, 1984.

INFORMATION THEORETIC COMPARISONS OF ORIGINAL AND TRANSFORMED DATA FROM LANDSAT MSS AND TM*+

William A. Malila
Environmental Research Institute of Michigan
Ann Arbor, MI 48107

ABSTRACT

A communications-theory approach is taken to analyze the dispersion and concentration of signal values in various data spaces, irrespective of any specific class memberships. Entropy, as defined by Shannon, is used to quantify information. Mutual information is used to measure the information represented by subsets of spectral variables. Examples of the concepts are presented. Several different comparisons of information content are made. These include comparisons of system design capacities, of data volumes occupied by agricultural data in the spaces defined by original bands and by transformed spectral (Tasseled Cap) variables, of the information contents of original bands and Tasseled Cap variables, and of the information contents of MSS and TM for the given agricultural data sets.

1. INTRODUCTION

With multispectral data sets from remote sensing systems, questions arise as to the relative merits of individual and groups of spectral bands and transformed spectral variables. Classification-based measures are frequently used for such comparisons, as are variance-based measures such as principal component analysis.

The first objective of the effort reported here was to develop a class-independent and non-parametric measure of the information content of multispectral data. The second objective was to use it to analyze and compare data from the two Landsat-4 sensors, the Multispectral Scanner System (MSS) and the Thematic Mapper (TM).

2. METHOD

A communications-theory approach is taken to analyze the dispersion and concentration of signal values in various data spaces. Entropy, as defined by Shannon, is used to quantify information. The process of selecting a subset of bands is viewed as the transmission of data through a lossy communication channel, and the mutual information between input and output is

*Presented at the Eighteenth International Symposium on Remote Sensing of Environment, Paris, France, October 1-5, 1984.

+This research was sponsored by the U.S. National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, MD, under Contract NAS5-27346.

used to measure information transfer, i.e., the information represented by the subset.

The alternative measure is applied to MSS and six-band TM data of two types. These are real Landsat-4 MSS and TM data acquired simultaneously from an agricultural scene in North Carolina, and data values synthesized from field-measured reflectance spectra of agricultural crops and soils using an atmospheric model. These data were used in prior comparisons of the spatial and spectral characteristics of Landsat TM and MSS data [1,2]. In the synthetic data, samples are primarily from vegetation at a variety of ground cover percentages, with many fewer examples of bare soil. All analyses of TM data are limited to the six reflective bands; the thermal band is not analyzed in this effort due to its coarser spatial resolution, its dependence on emissive rather than reflective characteristics of scene materials, and lack of a simulation data base.

Several different comparisons of information content are made. These include comparison of TM and MSS system-design information capacities, comparison of the data-space volumes spanned by the agricultural data in the spaces defined by original bands and by transformed spectral (Tasseled Cap) variables, comparison of the agricultural information content of original bands to that of transformed variables, and comparison of the agricultural information content of TM data to that of MSS.

2.1 INFORMATION MEASURE DERIVATION

2.1.1 Basic Concepts. Shannon defined self information, $I(x_i)$, as a measure of the information associated with knowing the occurrence of a signal state x_i which occurs with probability $P(x_i)$:

$$I(x_i) = \log_2\left(\frac{1}{P(x_i)}\right) = -\log_2 P(x_i) \qquad \text{(bits)} \tag{1}$$

The more rare the event, the greater is one's uncertainty about when it will occur and, consequently, the greater is the information conveyed when it is observed. Entropy, given the symbol H, is the value of self information when averaged over all N possible states of x:

$$H(x) = \sum_{i=1}^{N} P(x_i) \log_2 \frac{1}{P(x_i)} \tag{2}$$

With two variables, the use of joint and conditional probabilities is necessary:

$$H(x,y) = H(x) + H(y|x) \tag{3}$$

since

$$P(x,y) = P(x)P(y|x) \tag{4}$$

In computing the conditional entropy, the weighting assigned to each information term is the joint probability of the states involved, i.e.,

$$H(x|y) = \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} P(x_i,y_j) \log_2 \frac{1}{P(x_i|y_j)} \tag{5}$$

If we consider x to be the input to a communication channel and y to be the output, we can define the mutual information transferred between them,

ORIGINAL PAGE IS
OF POOR QUALITY

i.e., $I_M(x;y)$, as

$$I_M(x;y) = H(x) - H(x|y)$$

In words, the mutual information exchanged is the difference between $H(x)$, the information content of the input, and $H(x|y)$, the information loss or uncertainty about x when we are given the output y. When the total information is transferred, $H(x|y) = 0$ and $I_M(x;y) = H(x)$. At the other extreme, when y does not contain any information relatable to x, $H(x|y) = H(x)$ and therefore $I_M(x;y) = 0$, i.e., there is no mutual information.

Figure 1(a) presents a concise graphical summary of these quantities and their interrelationships. Note the special cases in Figure 1(b). Figures 2 and 3 present simple numerical examples that illustrate the concepts of entropy and joint entropy in a quantitative fashion. Note that entropy is at its maximum when all cells or states are equally likely. It can be reduced by decreasing the number of cells occupied, by having a non-uniform distribution or concentration of observations in the occupied cells, or by doing both.

2.1.2 Multispectral Extension. The above concepts can be extended to multispectral variables by letting the variables x and y become multidimensional vectors X and Y, with $X = (X_1, X_2, \ldots, X_{N_x})$ and $Y = (Y_1, Y_2, \ldots, Y_{N_y})$. Usually, $N_y \leq N_x$. The transformation achieved by the communication channel is used here in a general sense, to represent both simple selections of spectral band subsets and more complex transformations, such as the Tasseled Cap Transformation.

The following relationship was derived to compute entropy from counts of spectral cell populations (shown here for six variables):

$$H(X) = \underbrace{\log_2 N_{obs}}_{\substack{\text{Information}\\ \text{if each}\\ \text{observation}\\ \text{were in a}\\ \text{unique cell}}} - \underbrace{\left(\frac{1}{N_{obs}}\right) \sum_{ijklmn}\cdots\sum C_{ijklmn} \log_2 C_{ijklmn}}_{\substack{\text{Information loss due to concentration}\\ \text{of the observations into a subset of}\\ \text{cells}}} \tag{7}$$

where C_{ijklmn} is the count of occurrences in the cell having Level i in X_1, Level j in X_2, etc.,

and N_{obs} is the total number of observations in the data set being analyzed.

The entropy of X is expressed in Equation (7) as the difference between two terms. The first, $\log_2 N_{obs}$, is the maximum possible information associated with the given number of observations, i.e., the information that would be present if each observation were unique and occupied a unique cell in the signal space. The second term represents the information that is lost by any concentration of observations into a subset of cells.

2.1.3 Spectral Band Subsetting. The selection of subsets of spectral bands is a special case of the mutual information expression,

$$I_M(X;Y) = H(X) - H(X|Y)$$

where Y now is a subset, X', of the X variables, so

$$I_M(X;X') = H(X) - H(X|X')$$

Whenever a variable, say Xp, is retained, its conditional probability term becomes unity, its contribution to H(X| X') is reduced to zero, and its information content is retained as mutual information. Whenever a variable, say Xq, is eliminated, there is a loss of mutual information. This loss is represented by the conditional entropy term through all conditional probability components in which Xq occurs on the left-hand side of the conditional probability indicator line but not on the righthand (or given) side.

2.1.4 Spectral Transforms. Spectral transformations were obtained by applying the linear-combination Tasseled Cap (TASCAP) transformations to MSS [3] and TM [4] data. The principal TASCAP variables are Brightness and Greenness. Also, principal-component analysis was utilized to obtain a different set of spectral variables for one comparison.

3 RESULTS

3.1 SPECTRAL DATA VOLUMES

The diagram in Figure 4 helps describe the various terms used here to designate spectral data-space characteristics, while Table I quantifies many of the observed values. Figure 5 presents information measures for two of those quantities, as a function of the number of data variables. First, the system-design capacities of the Landsat-4 TM and MSS are presented, in terms of the number of bits transmitted to the ground and/or recorded on computer-compatible tapes (CCTs). For TM, the number of bits recorded on CCTs is the same as that transmitted (8 bits/channel). For MSS, however, the six-bit telemetered data are expanded to seven bits on the CCTs, with only n apparent gain of information. Nevertheless, many comparisons involving S will use seven-bit data since that is the form in which we received them. For some others, a degradation to six bits was performed before analysis. The greater information potential of the TM system design (reflective bands), as compared to the MSS system, is quantified as 48 vs 24 bits in telemetered data.

Figure 5 also portrays the "hypercube" volume or data-space volume spanned by TM and MSS data. These volumes are computed by summing the bit equivalents of the observed data-value ranges (max - min + 1) in each band being considered. Upon comparing the fractions of their total data-space volumes that are spanned by data from the agricultural scene, one observes that the TM data fall nine bits short of capacity while the MSS data fall approximately six bits short of capacity.

Actual data dispersion volumes (see Figure 4) were found to be substantially smaller than the hypercube volumes. Results for both real and synthetic data are shown in Figure 6 for MSS (7 bits/band; CCT) and Figure 7 for TM. (Note that these values for actual information are substantially smaller than reported elsewhere for similar comparisons in which the hypercube volumes are treated as the information content [5].) The data dispersion volumes in Figures 6 and 7 are measured by the entropies of the best variable combinations for the respective observation sets, and represent the information present in those sets. Most of the information is contained in the first two or three variables. The number of observations analyzed establishes a maximum limit on each entropy value. As shown earlier in Equation (7), the concentration of multiple observations (pixels) into individual spectral cells reduces the information content below the potential maximum. Table I shows very little tendency for TM pixels to do this, due to the very large system capacity, spectral diversity, and fine gradation of the TM bands. The MSS data show definite tendencies for multiple observations in spectral cells.

Table I shows that the TM data represent 3.3 bits more information than the MSS sensor data, with approximately two bits being associated with spatial resolution (pixel size and number) and the remainder with spectral bands and radiometric resolution. Since the synthetic data have the same number of observations for both TM and MSS, they can be considered to have equal spatial resolutions. Thus, the 2.2-bit difference must be due to their spectral and radiometric properties.

3.2 SPECTRAL TRANSFORMATIONS

Figure 8 compares the data-space volumes spanned by original and transformed versions of signals from the agricultural scene. It appears that a bit-rate reduction of about 3 bits/pixel could be achieved for this agricultural scene, without loss of information (see discussions of Figures 9 and 10), by transmitting values from the transformed variables instead of from the original bands. These differences might be greater for data sets with a broader range of scene amplitudes.

Figure 9 compares the agricultural information content of original and TASCAP variables from TM and MSS for the North Carolina scene. In each case, the best subset of each size was used. Here again, relatively little information is gained by the inclusion of more than three variables.

Figure 10 illustrates, for the synthetic MSS data set, the fact that the information content of original band values and two types of transformed variables are essentially identical. In addition to TASCAP variables, the information content of principal-component variables for this data set is also displayed. The equality of the complete sets of variables is in keeping with theoretical considerations of linear transformations.

3.3 SUBSETS OF VARIABLES

Mutual information values for the best and worst original-band subsets of each size are presented in Figure 11, to illustrate the range of information conveyed by various subsets of the variables. The differences are greatest among pairs of variables for both TM and MSS. Figure 12 is a similar comparison for TASCAP variables. In this case, we find an even greater disparity between best and worst combinations, due to the decreased information content of the last TASCAP variables.

3.4 DIMENSIONALITY

Figure 13 displays information measures computed for the first three Tasseled Cap components of TM and MSS data from the agricultural scene. (The MSS data were in CCT form at seven bits/band.) The first three components are individually quite similar for TM, but there is a substantial decrease (3.3 bits below Brightness) for the third component of MSS (Yellowness). This is consistent both with many investigators' experiences in finding MSS data of agricultural areas to be primarily two dimensional and with recent studies which have found a substantial amount of information in the TM Tasseled Cap Third Component [4]. Throughout this comparison, TM values are greater than the corresponding MSS values, for example the TM Brightness value is 6.7 bits compared to 5.8 bits for MSS.

When pairs of components are considered, we see substantial increases in total information, as would be expected with the addition of a second variable; the value for TM Brightness/Greenness is 4.8 bits greater than for Brightness alone, and the corresponding increase for MSS is 3.7 bits. However, differences do appear between MSS and TM. Whereas the value of the Brightness/Greenness pair for MSS is substantially greater than the other two (approximately two bits greater than Greenness/Third Component), there again

is relatively little difference (less than 0.4 bits) among the three pairings from TM data, pointing to a higher dimensionality in TM.

Three components captured the vast majority of information for both systems. However, the fact that the gain in going from two to three components was nearly as large for MSS (1.25 bits) as for TM (1.7 bits) was somewhat surprising in view of the previously discussed two-dimensional character of MSS data. Furthermore, principal-component analysis of MSS data showed nearly total representation of variance by the first two components. The MSS gain likely is due to the Brightness/Greenness plane having a thickness of several counts in the third direction, even though this third component was uncorrelated with the others. The observed values also indicate that differences do exist among these various measures of multispectral signal properties. The TM data pattern also may be somewhat planar in three space, although not aligned as well with any component axis; correlations with the Third Component were -0.69 for Brightness and 0.36 for Greenness in this data set. None of these observations should diminish the utility [3,4] of Tasseled Cap transforms for physical interpretation of data values and agricultural scene characteristics.

3.5 NOISE

Noise in multispectral data was not considered explicitly in the results presented herein. Sensor noise effects certainly were present in the real Landsat data and natural variations of crop observations were present in both the real and synthetic data. Noise can add variance to signals and increase the number of spectral cells occupied (above that for no noise), thereby creating an apparent information content greater than the true information content of ideal, noiseless signals. One might address such effects by reducing the number of discrete levels present in data sets by applying an appropriate quantization factor (greater than unity) to each band and computing the reduced information content.

4. SUMMARY DISCUSSION

An information-theoretic measure was defined and applied to Landsat multispectral data, both real and synthetic. Examples of the basic concepts also were generated. The measure does quantify signal dispersion patterns, independently of class membership and distributional assumptions. It also provides an alternate method (to classification) of measuring the extent to which subsets of bands or transformed variables represent the total pattern. In planning analyses and interpreting results, however, analysts should insure that data sets being analyzed are representative of the problems under consideration.

A number of observations were made from this initial study. The system-design information capacity of TM is much greater than that of MSS. The potential information capacities and the signal "hypercube" volumes of agricultural data were much larger than the information actually represented by signal dispersion patterns in the sets of data values analyzed. Tasseled Cap transformations preserved the information in original bands and offered a modest savings in bits over those o 'ginal bands, a fact which might be useful in data compression approaches. There were extremely few multiple occurrences of spectral observations in the TM data set, but a reasonably high number for the MSS data, another indication of TM's finer partitioning of spectral space. For the "best" combinations of variables, entropy magnitudes were more a function of the number of variables than of the type of variables (original bands or tranformed). TM had greater entropy values for Brightness and Brightness/Greenness than did MSS. Information in the Tasseled Cap Third Component of TM was much greater than that of MSS, both by itself and in combination with Brightness or Greenness, confirming TM's greater dimensionality.

In future studies, it is recommended that additional data sets be analyzed, both with larger sample sizes and with varied scene contents; effects of other transformations might also be examined. Noise effects should be investigated through use of quantization factors to degrade radiometric resolutions. It may also be fruitful to investigate approaches to incorporate class membership into information-theoretic measures of multispectral information content.

REFERENCES

1. Malila, W.A., M.D. Metzler, D.P. Rice and E.P. Crist, "Characterization of Landsat-4 MSS and TM Digital Image Data", IEEE Transactions on Geoscience and Remote Sensing, pp.177-191, May 1984.

2. Crist, E.P., "Comparison of Coincident Landsat-4 MSS and TM Data Over an Agricultural Region", Technical Papers of the 50th Annual Meeting of the American Society of Photogrammetry, Vol.2, pp.508-517, 1984.

3. Kauth, R.J. and G.S. Thomas, "The Tasseled-Cap -- A Graphic Description of the Spectral-Temporal Development of Agricultural Crops as Seen by Landsat", Symposium on Machine Processing of Remotely Sensed Data, June 29 - July 1, 1976, Purdue University, LARS, W.Lafayette, IN, 1976.

4. Crist, E.P. and R.C. Cicone, "A Physically-based Transformation of Thematic Mapper Data - The TM Tasseled Cap", IEEE Transactions on Geoscience and Remote Sensing, pp.256-263, May 1984.

5. Price, J.C., "Comparison of the Information Content of Data from the Landsat-4 Thematic Mapper and Multispectral Scanner", IEEE Transactions on Geoscience and Remote Sensing, pp.272-281, May 1984.

Table I. Information Comparison for MSS and Six-Band TM Data Sets

A. VALUES FOR REAL AGRICULTURAL DATA (N. CAROLINA SCENE)

		MSS NUMBER	MSS BITS	SIX-BAND TM NUMBER	SIX-BAND TM BITS	TM GAIN (BITS)
SYSTEM CAPACITY:	SENSOR	0.17×10^8	24	0.28×10^{15}	48	24
	CCT	0.27×10^9	28	0.28×10^{15}	48	20
HYPERCUBE VOL.:	SENSOR	0.44×10^6	18.7	0.43×10^{12}	38.6	
	CCT	0.32×10^7	21.6	0.43×10^{12}	38.6	
DATA DISPERSION PATTERN						
MSS CCT: 7 Bits per Band	· # Observations, H_{max}	3,468	11.8	13,015	13.7	$\boxed{1.91}$ (Spatial)
	· # Unique cells	2,898	-	12,903	-	
	· Entropy, H	-	11.4	-	13.7	$\boxed{2.27}$ (Total)
	· Loss due to spectro-radiometric concentration	-	0.38	-	0.02	$\boxed{0.36}$ (Spec/Radiom)
MSS Sensor 6 Bits per Band	# Unique cells	1,730	-	12,903	-	
	Entropy, H	-	10.3	-	13.7	$\boxed{3.34}$ (Total)
	· Loss due to spectro-radiometric concentration	-	1.45	-	0.02	$\boxed{1.43}$ (Spec/Radiom)

B. VALUES FOR SYNTHETIC AGRICULTURAL DATA (Assumes equal spatial resolution)

	MSS NUMBER	MSS BITS	SIX-BAND TM NUMBER	SIX-BAND TM BITS	TM GAIN (BITS)
"SYSTEM" CAPACITY (MSS: 6 Bits/Band)	0.17×10^8	24	0.28×10^{15}	48	24
OBSERVED HYPERCUBE VOLUME	0.10×10^7	20	0.99×10^{12}	40	20
DATA DISPERSION PATTERN:					
· # Observations; H_{max}	2,276	11.15	2,276	11.15	
· # Unique cells	817	-	2,260	-	
· Entropy, H	-	8.94	-	11.14	$\boxed{2.20}$
· Loss due to spectro-radiometric concentration	-	2.21	-	0.014	(Spec/Radiom)*

*(TM gain over seven-bit MSS data was one bit.)

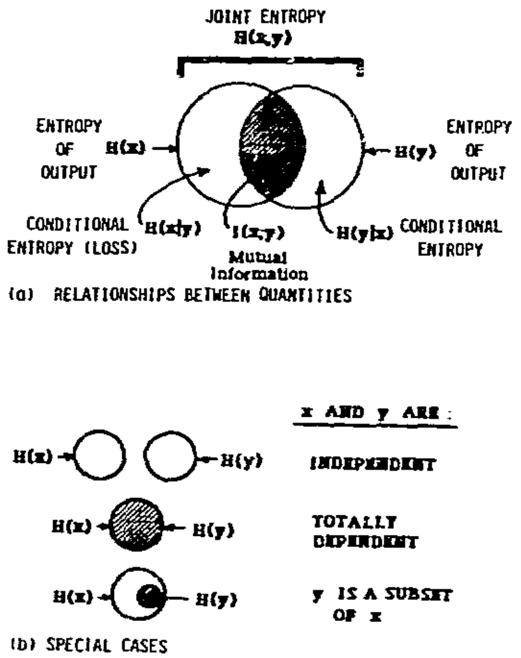

Figure 1. Basic Information Concepts

GIVEN:
- SEVERAL FOUR-SIDED SOLIDS.
- EACH TOSSED EIGHT TIMES AND "DOWN" SIDES TALLIED.
- EXPERIMENT REPEATED SEVERAL TIMES TO OBTAIN AVERAGE OCCURRENCE RATES (SHOWN BELOW).

PROBLEM:
- COMPUTE AND COMPARE ENTROPIES

COUNT FOR SIDE:

	S1	S2	S3	S4
SOLID A:	2	2	2	2

$P(x_i)$: 1/4 1/4 1/4 1/4

$Log_2(1/P_i)$: 2 2 2 2

$H(x) = \sum P_i Log_2(1/P_i) = 1/2 + 1/2 + 1/2 + 1/2 = 2$ Bits — MAXIMUM VALUE FOR FOUR STATES (Equally Likely)

	S1	S2	S3	S4
SOLID B:	1	2	4	1

$P(x_i)$: 1/8 1/4 1/2 1/8

$Log_2(1/P_i)$: 3 2 1 3

$H(x) = \sum P_i Log_2(1/P_i) = 3/8 + 1/2 + 1/2 + 3/8 = 7/4$ Bits

	S1	S2	S3	S4
SOLID C:	2	2	4	0

(P_i): 1/4 1/4 1/2 0

$Log_2(1/P_i)$: 2 2 1 -

$H(x) = 1/2 + 1/2 + 1/2 + 0 = 3/2$ Bits

	S1	S2	S3	S4
SOLID D:	0	4	4	0

P_i: 0 1/2 1/2 0

$Log_2(1/P_i)$: - 1 1 -

$H(x) = 0 + 1/2 + 1/2 + 0 = 1$ Bit — MAXIMUM VALUE FOR TWO STATES (Equally Likely)

	S1	S2	S3	S4
SOLID E:	0	8	0	0

P_i: 0 1 0 0

$Log_2(1/P_i)$: - 0 - -

$H(x) = 0 + 0 + 0 + 0 = 0$ Bits — NO UNCERTAINTY, THEREFORE, NO INFORMATION VALUE

Figure 2. Entropy Examples

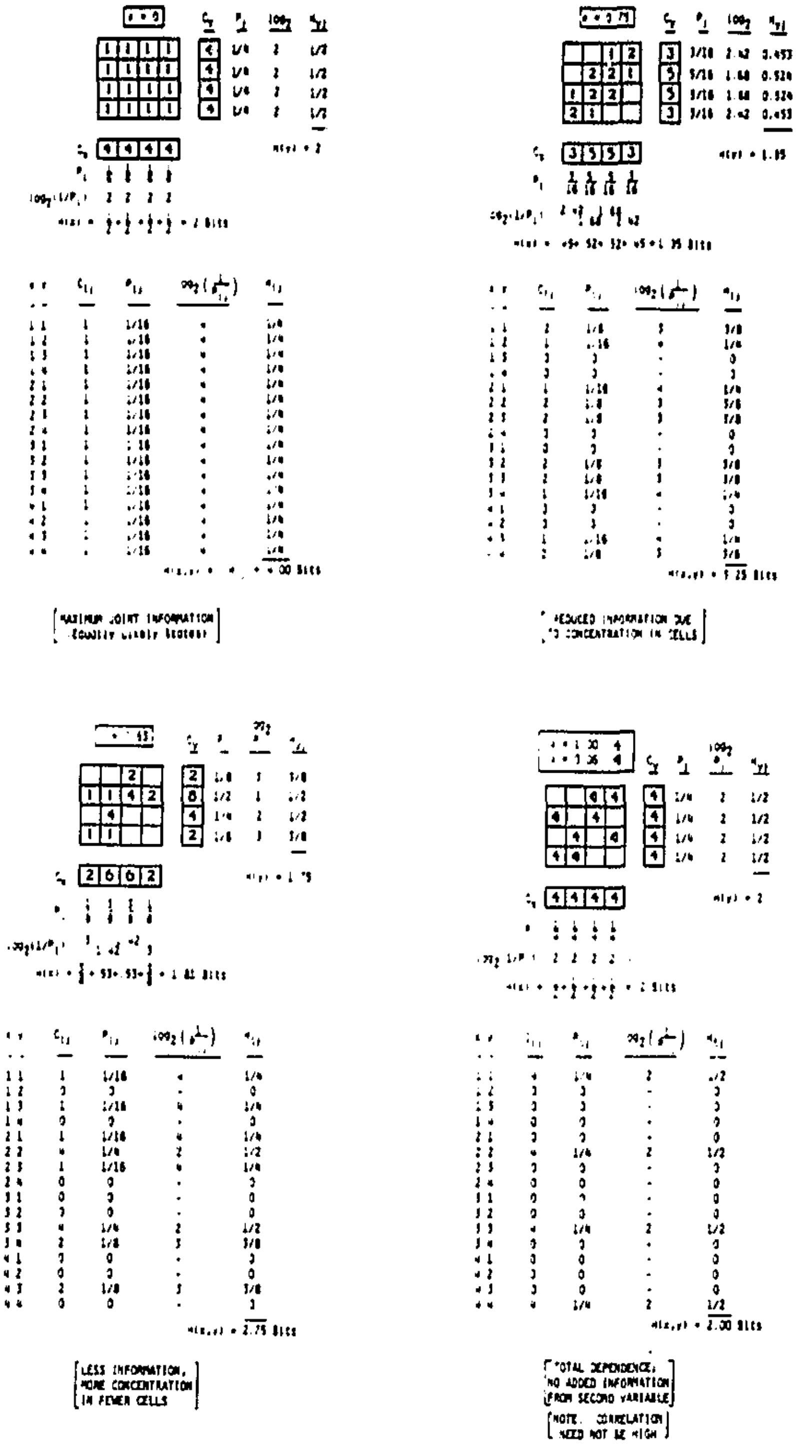

Figure 3. Joint Entropy Examples

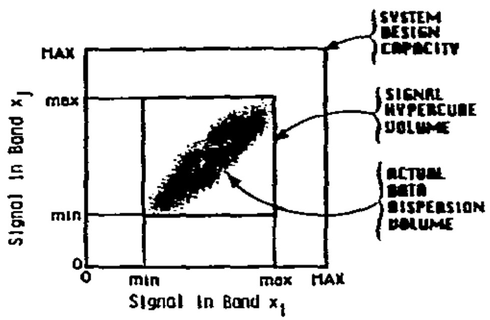

Figure 4. Illustration of Various Multispectral Data Volumes

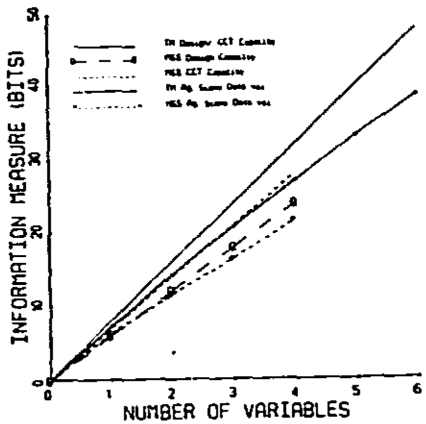

Figure 5. Comparison of TM & MSS Information Capacities

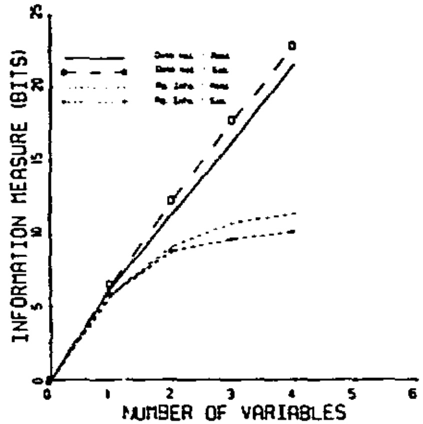

Figure 6. Comparison of Real & Simulated Agricultural Data Sets: MSS (7 bits/band)

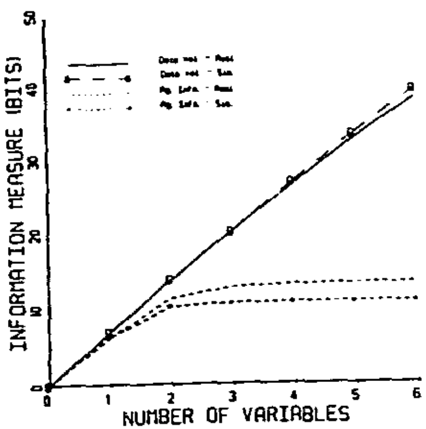

Figure 7. Comparison of Real & Simulated Agricultural Data Set: TM

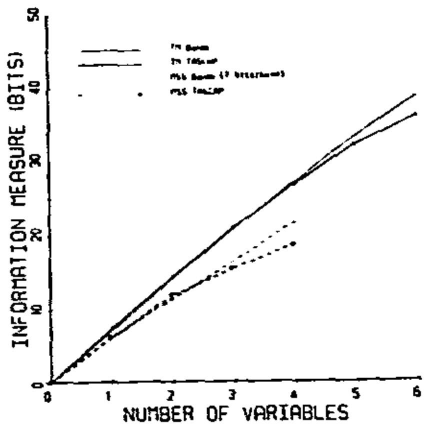

Figure 8. Comparison of Data Volumes: Real Agricultural Data

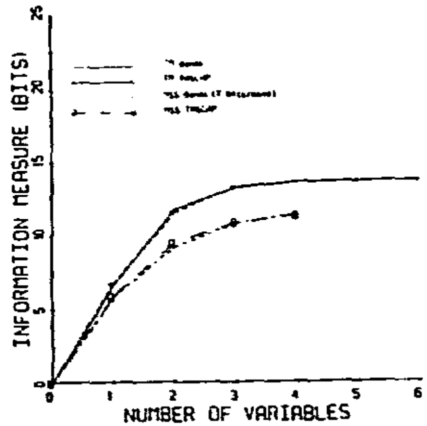

Figure 9. Comparison of Agricultural Information Content Real Data

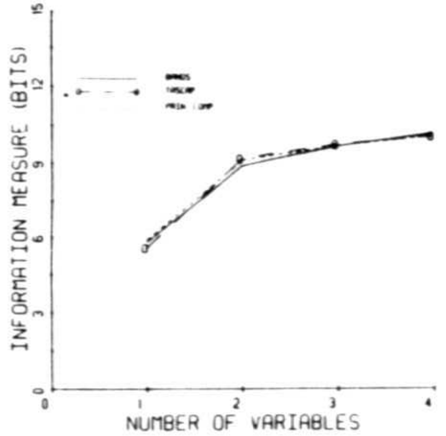

Figure 10. Information Comparison of MSS Band, TASCAP & Principal-Component Variables (Simulated Ag. Data)

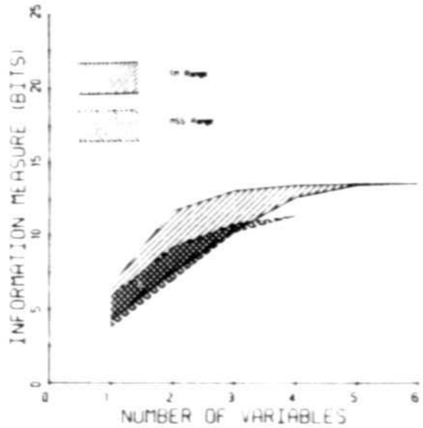

Figure 11. Range of Information In Subsets of Bands: Real Agr. Data

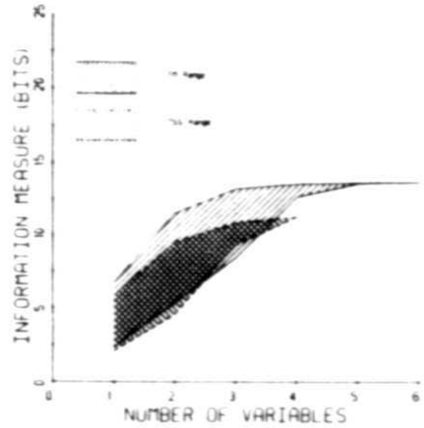

Figure 12. Range of Information in Subsets of TASCAP Variables: Real Agr. Data

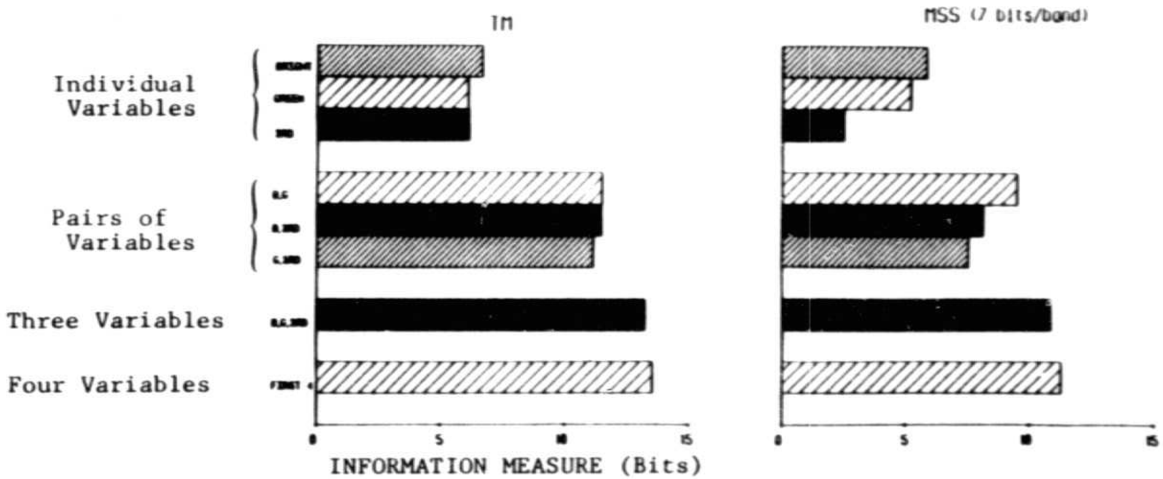

Figure 13. TM-MSS TASCAP Comparisons: Real Agricultural Data